A thought experiment in many worlds

Shan Gao
Research Center for Philosophy of Science and Technology,
Shanxi University, Taiyuan 030006, P. R. China
E-mail: gaoshan2017@sxu.edu.cn.

February 17, 2020

Abstract

The many-worlds interpretation of quantum mechanics (MWI) is based on three key assumptions: (1) the completeness of the physical description by means of the wave function, (2) the linearity of the dynamics for the wave function, and (3) multiplicity. In this paper, I propose a new thought experiment in which a post-measurement superposition undergoes no net change while individual branches do change under certain unitary time evolution. Moreover, I argue that MWI gives contradictory predictions for this experiment. In order to avoid the contradiction and save many worlds, it seems that we must drop one or both of the first two assumptions.

1 Introduction

The many-worlds interpretation of quantum mechanics (MWI) assumes that the wave function of a physical system is a complete description of the system, and the wave function always evolves in accord with the linear Schrödinger equation. In order to solve the measurement problem, MWI further assumes that after a measurement with many possible results there appear many equally real worlds, in each of which there is a measuring device which obtains a definite result (Everett, 1957; DeWitt and Graham, 1973; Barrett, 1999; Wallace, 2012; Vaidman, 2018).[1] In this paper, I will propose a new thought experiment and argue that MWI gives contradictory predictions for this experiment.

[1]This assumption is supported by an extensive analysis of decoherence and emergence in the modern formulations of MWI (see, e.g. Wallace, 2012).

2 A thought experiment

Consider a usual situation where a measuring device M interacts with a measured system S. When the state of S is $|1\rangle_S$, the state of M changes and it obtains a result 1:

$$|1\rangle_S |ready\rangle_M \rightarrow |1\rangle_S |1\rangle_M . \tag{1}$$

When the state of S is $|2\rangle_S$, the state of M changes and it obtains a result 2:

$$|2\rangle_S |ready\rangle_M \rightarrow |2\rangle_S |2\rangle_M . \tag{2}$$

According to MWI, there is no world branching and there is still one measuring device, namely the original one, after the above measurement.

Now suppose the measuring device M interacts with the system S being in a superposed state $|1\rangle_S + |2\rangle_S$. For simplicity I omit the nomalization factor $1/\sqrt{2}$. By the linearity of the dynamics for the wave function, the state of the composite system after the interaction will evolve into the following superposition:

$$|1\rangle_S |1\rangle_M + |2\rangle_S |2\rangle_M . \tag{3}$$

According to the multiplicity assumption of MWI, this post-measurement superposition corresponds to two (or two sets of) worlds, in each of which there is a measuring device which has a definite state, obtaining a result 1 or obtaining a result 2. We may denote these two worlds as W_{11} and W_{22}. Here I omit the environment terms in the evolution, which, in a more complete form, should be $|1\rangle_S |1\rangle_M |1\rangle_E + |2\rangle_S |2\rangle_M |2\rangle_E$. This does not influence my following analysis when viewing the whole composite system including the environment as an isolated system.

Consider a unitary time evolution operator U_N which changes $|1\rangle_S |1\rangle_M$ to $|2\rangle_S |2\rangle_M$ and $|2\rangle_S |2\rangle_M$ to $|1\rangle_S |1\rangle_M$.[2] Then by the linearity of the dynamics, the time evolution of the above post-measurement state under U_N is

$$U_N(|1\rangle_S |1\rangle_M + |2\rangle_S |2\rangle_M) = |2\rangle_S |2\rangle_M + |1\rangle_S |1\rangle_M . \tag{4}$$

It can be seen that after the unitary time evolution U_N the post-measurement state does not change. Then, according to the multiplicity assumption of MWI, the state after U_N still corresponds to two worlds, W_{11} and W_{22}, in

[2]For a Hilbert space with dimension greater than two, the swap operator U_N can be accomplished in many ways, such as with a 180 degree rotation about the ray halfway between the two state vectors. I thank ... for this useful comment. Note that a similar thought experiment involving the swap operator U_N was first proposed and discussed by Gao (2019).

each of which there is a measuring device which has a definite state, either obtaining a result 1 or obtaining a result 2.

An interesting question now arises: what is the corresponding relation between the worlds after U_N and the worlds before U_N? It can be expected that the corresponding relation depends on how U_N is accomplished. Let us see a form of U_N which satisfies the criteria for decoherence so that the histories can have consistent transition probabilities assigned. Suppose the dimensions of the Hilbert spaces of both the system and the measuring device are greater than two. Then we also have the bases $|1\rangle_S |2\rangle_M$ and $|2\rangle_S |1\rangle_M$ besides $|1\rangle_S |1\rangle_M$ and $|2\rangle_S |2\rangle_M$. Moreover, suppose U_N first changes $|1\rangle_S |1\rangle_M$ to $|2\rangle_S |1\rangle_M$ (during which $|1\rangle_M$ keeps unchanged) and then changes $|2\rangle_S |1\rangle_M$ to $|2\rangle_S |2\rangle_M$ (during which $|2\rangle_S$ keeps unchanged).[3] Similarly, U_N first changes $|2\rangle_S |2\rangle_M$ to $|1\rangle_S |2\rangle_M$ and then changes $|1\rangle_S |2\rangle_M$ to $|1\rangle_S |1\rangle_M$, namely

$$|1\rangle_S |1\rangle_M \rightarrow |2\rangle_S |1\rangle_M \rightarrow |2\rangle_S |2\rangle_M \tag{5}$$

and

$$|2\rangle_S |2\rangle_M \rightarrow |1\rangle_S |2\rangle_M \rightarrow |1\rangle_S |1\rangle_M \,. \tag{6}$$

Then, at any time t during the process of U_N there will be in general a superposition of four branches:

$$c_{11}(t) |1\rangle_S |1\rangle_M + c_{12}(t) |1\rangle_S |2\rangle_M + c_{21}(t) |2\rangle_S |1\rangle_M + c_{22}(t) |2\rangle_S |2\rangle_M \,. \tag{7}$$

Since these branches are decoherent, U_N satisfies the criteria for decoherence; there will be in general four worlds during the process of U_N: W_{11}, W_{12}, W_{21}, and W_{22}. Moreover, the histories (5) and (6) can have consistent transition probabilities assigned, and the transition probabilities of both histories are one.[4] This means that after U_N, W_{11} changes to W_{22} through W_{21}, and W_{22} changes to W_{11} through W_{12}.

3 A contradiction

Now we can prove that the combination of the three key assumptions of MWI leads to a contradiction in the above thought experiment.

According to the assumption of multiplicity, the post-measurement superposition $|1\rangle_S |1\rangle_M + |2\rangle_S |2\rangle_M$ corresponds to two worlds, W_{11} and W_{22}, in each of which there is a measuring device which has a definite state, either obtaining a result 1 or obtaining a result 2. According to the linearity of the dynamics for the wave function, this superposition does not change

[3] Note that U_N is essentially continuous and needs a finite time to be accomplished.

[4] I thank ... for this helpful comment.

after the time evolution U_N defined by (1) and (2). According to the completeness of the physical description by means of the wave function, since this superposition does not change after U_N, the *complete* physical state of the composite system being in this superposition does not change after U_N. This means that every aspect of the state of reality of the composite system does not change after U_N, and in particular, the state of each real world the system comprises does not change. If anything physical changes after U_N, then the physical description by means of the wave function will be not complete, since the wave function does not change after U_N.

On the other hand, after a certain U_N defined by (5) and (6), the state of each world does change; W_{11} changes to W_{22}, and W_{22} changes to W_{11}. In particular, the state of the measuring device M in each world changes, either from obtaining result 1 to obtaining result 2 or from obtaining result 2 to obtaining result 1. Thus, we prove that the three key assumptions of MWI leads to a contradiction in the above thought experiment.

One may think that it is possible for the complete state of a system to be unchanged even if the states of some parts of the system change. For example, consider a simple non-quantum analogy.[5] Imagine two similar spherical bodies orbiting one another in a universe where nothing else is changing. As the bodies orbit, they change colors. At an initial instant, there are a green body on the left and a red one on the right. After each completes a half rotation and they have swapped positions, there are again a green body on the left and a red one on the right (as the bodies have changed colors over this period). Then, one may conclude that the state of the universe is the same, but the state of each body has changed, since each has moved and changed color.

However, it can be seen that this example does not show that the state of the universe may keep unchanged even if the states of some parts of the universe change. Here is the reason. If the two bodies are identical except color, then after the transformation, the physical states of the two bodies have not changed, and the state of the universe has not changed either.[6] On the other hand, if the two bodies are not identical besides color, such as possessing different properties denoted by 1 and 2, then after the transformation the state of the universe have changed. Suppose at the initial instant the green body on the left is body 1 and the red body on the right is body 2. Then after the transformation, body 1 is on the right and body 2 is on the left, and thus both the states of the two bodies and the state of the universe have changed. Therefore, in either case, the above example does not show that the complete state of a system may keep unchanged even if the states of some parts of the system change.

[5] I thank ... for this interesting example.

[6] During the transforming process, the physical states of the two bodies do change, but the state of the universe also changes.

4 Further discussion

What about other U_Ns which do not satisfy the criteria for decoherence? Can we also prove the existence of a contradiction for these U_Ns? I think the answer may be positive.

First, since the post-measurement superposition before and after U_N are exactly the same, if the superposition before U_N corresponds to two worlds, then the superposition after U_N will also correspond to two worlds. This is true independently of the unitary time evolution U_N and whether U_N violates the criteria for decoherence.

Next, if the whole superposition represents the *complete* physical state and it does not change, then the underlying state of reality will not change and everything in the physical world will not change, including those emergent high-level objects. In this case, if individual branches do change (when the whole superposition undergoes no change), then they cannot represent something physical such as real worlds. On the other hand, if something physical changes and the whole wave function does not change, then the wave function cannot represent the *complete* physical state. For example, if individual branches represent real worlds as MWI assumes, then when the whole wave function undergoes no change while these worlds do not keep unchanged, the wave function cannot represent the complete physical state.

Now, if U_N does not satisfy the criteria for decoherence, then the histories of the initial two worlds do not have unique transition probabilities assigned (without resorting to auxiliary rules). For example, a certain form of U_N may bring about interference, and the initial two worlds will merge and two new worlds will appear during the time evolution.[7] In this case, MWI does not predict that the state of each world keeps unchanged after U_N. Then, if the completeness of the physical description by means of the wave function requires that each world should not change after U_N, then there will be still a contradiction.

Finally, it can be argued that the state of each world cannot keep unchanged after U_N, and thus the contradiction seems inevitable. Consider a more general post-measurement superposition: $\alpha |1\rangle_S |1\rangle_M + \beta |2\rangle_S |2\rangle_M$. Its time evolution under U_N is:

$$\alpha |1\rangle_S |1\rangle_M + \beta |2\rangle_S |2\rangle_M \rightarrow \alpha |2\rangle_S |2\rangle_M + \beta |1\rangle_S |1\rangle_M . \tag{8}$$

Worlds and their evolution are arguably independent of the values of α and β. If the state of each world keeps unchanged after U_N, then the term $\alpha |1\rangle_S |1\rangle_M$ will evolve to $\beta |1\rangle_S |1\rangle_M$, and the term $\beta |2\rangle_S |2\rangle_M$ will evolve to $\alpha |2\rangle_S |2\rangle_M$. But this violates the linearity of the dynamics for the wave function. In particular, when $\alpha = 1$ and $\beta = 0$, it means that $|1\rangle_S |1\rangle_M$ will evolve to 0, and 0 will evolve to $|2\rangle_S |2\rangle_M$; this is impossible.

[7] I thank ... for raising this insightful objection.

There are two possible ways to avoid the above contradiction. The first way is to deny that after the evolution the state of the composite system has not changed. This requires that the wave function of a system is not a complete description of the state of the system, and additional variables are needed to introduce to describe the complete state. In other words, the assumption of the completeness of the physical description by means of the wave function should be dropped. The second way is to deny that after the evolution the states of the worlds have changed. This is possible when there is only one world or the linearity of the dynamics is violated. In other words, the assumption of multiplicity or the assumption of the linearity of the dynamics for the wave function should be dropped.

5 Conclusion

In this paper, I propose a new thought experiment in which a post-measurement superposition undergoes no net change while individual branches do change under certain unitary time evolution. I argue that MWI gives contradictory predictions for this experiment. The contradiction results from the combination of three key assumptions of MWI: (1) the completeness of the physical description by means of the wave function, (2) the linearity of the dynamics for the wave function, and (3) multiplicity. In order to avoid the contradiction and save many worlds, it seems that we must drop one or both of the first two assumptions, although this is against the spirit of modern formulations of MWI.

Acknowledgments

I wish to thank Jeff Barrett, Sean Carroll, David Deutsch, Meir Hemmo, Louis Marchildon, Don Page, Simon Saunders, Paul Tappenden, Lev Vaidman, David Wallace, and Kai-Ning Wang for helpful discussion. This work is supported by the National Social Science Foundation of China (Grant No. 16BZX021).

References

[1] Barrett, J. A. (1999). The Quantum Mechanics of Minds and Worlds. Oxford: Oxford University Press.

[2] DeWitt, B. S. and N. Graham (eds.). (1973). The Many-Worlds Interpretation of Quantum Mechanics. Princeton: Princeton University Press.

[3] Everett, H. (1957). 'Relative state' formulation of quantum mechanics. Rev. Mod. Phys. 29, 454-462.

[4] Gao, S. (2019). The measurement problem revisited. Synthese, 196, 299-311.

[5] Vaidman, L. (2018). Many-Worlds Interpretation of Quantum Mechanics, The Stanford Encyclopedia of Philosophy (Fall 2018 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/archives/fall2018 /entries/qm-manyworlds/.

[6] Wallace, D. (2012). The Emergent Multiverse: Quantum Theory according to the Everett Interpretation. Oxford: Oxford University Press.